Biochimica et Biophysica Acta 1814 (2011) 1870–1879

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Prediction of residues involved in inhibitor specificity in the dihydrofolate reductase family

N.M. Goodey [a,*], K.G. Herbert [b], S.M. Hall [a], K.C. Bagley [a]

[a] Montclair State University, Department of Chemistry and Biochemistry, 1 Normal Ave., Montclair, NJ 07043, USA
[b] Montclair State University, Department of Computer Science, 1 Normal Ave., Montclair, NJ 07043, USA

ARTICLE INFO

Article history:
Received 25 March 2011
Received in revised form 29 July 2011
Accepted 1 August 2011
Available online 5 August 2011

Keywords:
Drug-target interaction
Target prediction
Dihydrofolate reductase
Allosteric interaction
Inhibitor specificity
Specificity determinant

ABSTRACT

Dihydrofolate reductase (DHFR) is of significant recent interest as a target for drugs against parasitic and opportunistic infections. Understanding factors which influence DHFR homolog inhibitor specificity is critical for the design of compounds that selectively target DHFRs from pathogenic organisms over the human homolog. This paper presents a novel approach for predicting residues involved in ligand discrimination in a protein family using DHFR as a model system. In this approach, the relationship between inhibitor specificity and amino acid composition for sets of protein homolog pairs is examined. Similar inhibitor specificity profiles correlate with increased sequence homology at specific alignment positions. Residue positions that exhibit the strongest correlations are predicted as specificity determinants. Correlation analysis requires a quantitative measure of similarity in inhibitor specificity (S_{lig}) for a pair of homologs. To this end, a method of calculating S_{lig} values using K_I values for the two homologs against a set of inhibitors as input was developed. Correlation analysis of S_{lig} values to amino acid sequence similarity scores – obtained via multiple sequence alignments – was performed for individual residue alignment positions and sets of residues on 13 DHFRs. Eighteen alignment positions were identified with a strong correlation of S_{lig} to sequence similarity. Of these, three lie in the active site; four are located proximal to the active site, four are clustered together in the adenosine binding domain and five on the βFβG loop. The validity of the method is supported by agreement between experimental findings and current predictions involving active site residues.

1. Introduction

Dihydrofolate reductase (DHFR), a ubiquitous enzyme involved in major biological processes including DNA synthesis, is an essential enzyme for life. As such, DHFR constitutes a credible drug target in the treatment of many diseases, including cancer and microbial infections [1–3]. There is recent and significant interest in DHFR as a target for drugs against infectious diseases [4,5]. These include parasitic infections caused by protozoa *Leishmania major (Lm), Trypanosoma cruzi (Tc)* and *brucei (Tb),* and *Plasmodium falciparum (Pf)*) and opportunistic infections caused by *Pneumocystis carinii (Pc)* and *Toxoplasma gondii (Tg)* that are often fatal in patients with immune deficiencies caused by AIDS or bacterial infections [8–18]. Drug discovery aimed at DHFRs from these and other pathogenic organisms hold promise because DHFR is an established target in the treatment of cancer and other infections. Well known examples of DHFR inhibitors include the antibiotic trimethoprim (TMP), the protozoal inhibitors pyrimethamine (PYR), cycloguanil (CYC), and WR99210 and the chemotherapeutic agents methotrexate (MTX), trimetrexate (TMQ), and pemetrexed (LY231514) [6]. Raltitrexed (ZD1694) is a moderate inhibitor of DHFRs from several species but is primarily known as a thymidylate synthase inhibitor. Some inhibitors such as TMQ are known inhibitors for mammalian DHFRs, some for bacterial (e.g. TMP) and some for protozoal species (CYC, WR99210, and PYR), although many bind DHFRs from all types of organisms, albeit with different affinities.

Several known, highly conserved active site residues are involved in ligand binding in the DHFR family; many have been identified through the sequencing of drug resistant cell lines. Mutations at these positions (cross-species hot spots) tend to result in broad-based drug resistance in DHFRs from various species, indicating that these residues are binding to common features of DHFR inhibitors. For example, residue Phe34 (residue numbering according to human DHFR) is highly conserved within the DHFR family and mutations at this position result in low affinities for a variety of DHFR inhibitors in mammalian, bacterial, and protozoal DHFRs suggesting an important interaction between the

Abbreviations: DHFR, Dihydrofolate reductase; *Hs, H. sapiens*; *Ec, E. coli*; *Mt, M. tuberculosis*; *Tb, T. brucei*; *Tg, T. gondii*; *Lm, L. major*; *Tc, T. cruzi*; *Pf, P. falciparum*; *Pv, P. vivax*; *Rn, R. norvegicus*; *Pc, P. carnii*; *Lc, L. casei*; *Ma, M. avium*; MTX, Methotrexate; TMP, Trimethoprim; ZD1694, Raltitrexed; PTX, Piritrexim; LY231514, Pemetrexed; PYR, Pyrimethamine; MBP, Methylbenzoprim; CYC, Cycloguanil; TMQ, Trimetrexate; EPM, Epiroprim; CB3717, PDDF

* Corresponding author. Tel.: +1 973 655 3410; fax: +1 973 655 7222.
E-mail addresses: goodeyn@mail.montclair.edu (N.M. Goodey), herbertk@mail.montclair.edu (K.G. Herbert), halls6@mail.montclair.edu (S.M. Hall), bagleyk1@mail.montclair.edu (K.C. Bagley).

1570-9639/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2011.08.001

phenyl ring of Phe34 and a common feature shared by DHFR ligands [7,8]. DHFR residue 115 (*Hs* numbering) is another example of a position where side chain replacements results in decreased affinity for inhibitors in the multiple species that have been studied [9]. Such residues may be more important in ligand binding in general than ligand discrimination. On the other hand, there are other positions that define species-specific inhibition. For example, the residue position 31 (F31 in *Hs*) has been found to be important for ligand specificity in the DHFR family [10,11]. Achieving species specificity, and thus reducing toxic effects in the human host during treatment of infections, is a key goal in drug discovery in the DHFR family [12,13].

Identification of residues involved in determining the relative effects of different inhibitors on an enzyme's activity, the *inhibitor specificity profile,* is a key element in the rational design of selective small molecule inhibitors, protein engineering, and the prediction of resistance mutations [14]. Traditionally, residues involved in ligand binding are predicted based on interaction distance by visually inspecting X-ray or NMR 3D ligand bound structures. This approach is limited thought because it does not usually reveal which residues are responsible for a protein's inhibitor *specificity profile* or allow for predictions about the roles of residues distal to the ligand binding site [15]. More recently, phylogenetics-based drug discovery has examined the homology relations between proteins in certain families to predict protein–ligand interactions [16–25]. A general study of α-helical proteins showed that the chemical similarity between ligands is higher for proteins with similar sequences [26]. A correlation between the likelihood that two G-protein coupled receptors bind the same ligand and their sequence similarity has also been shown, no relationship was observed for a group of kinases [21].

In this paper, we report a novel phylogenetics-based method for predicting residues involved in inhibitor specificity in the DHFR family. In this approach, the relationship between inhibitor specificity and amino acid composition for a set of protein homolog pairs is evaluated. For specific alignment positions within a homologous protein family, similar inhibitor specificity profiles are shown to correlate with increased local sequence homology. Residue positions that exhibit the strongest correlations are predicted to be inhibitor specificity determinants. Correlation analysis requires a quantitative measure of similarity in inhibitor specificity (S_{lig}) for a pair of homologs. To this end, we established a method of calculating S_{lig} values for pairs of homologs using K_I values for the two homologs against a set of inhibitors as inputs. A database of K_I values for 13 DHFRs against multiple inhibitors was created and used to calculate S_{lig} values for all possible DHFR pairs for which sufficient data were available. The DHFRs were aligned and sequence similarity scores were determined for all possible pairs between the 13 DHFRs. Correlation analysis of S_{lig} values to amino acid sequence similarity scores was performed for individual residue alignment positions. The positions with the highest correlations were grouped and subjected to a correlation analysis where multiple residues positions were considered. A strong correlation between S_{lig} and sequence similarity was observed for set residues. The validity of the method is supported by identification of two active site residues (Phe31 and Ans61, *Hs* numbering) that have previously been experimentally shown to serve as specificity determinants.

2. Materials and methods

2.1. Creation of DHFR homolog-ligand database

A DHFR homolog-ligand database was created with data for 13 DHFR homologs. The 13 species included in the database are *Homo sapiens (Hs),Escherichia coli (Ec), Mycobacterium tuberculosis (Mt), Trypanosoma brucei (Tb), T. gondii (Tg), L. major (Lm), T. cruzi (Tc), P. falciparum (Pf), Plasmodium vivax (Pv), Rattus norvegicus (Rn), P. carinii (Pc), Lactobacillus casei (Lc) and Mycobacterium avium (Ma)*. This database contains K_I values for 12 different ligands (Methotrexate (MTX), Trimethoprim (TMP), Raltitrexed (ZD1694), Piritrexim (PTX), Pemetrexed (LY231514), Pyrimethamine (PYR), Methylbenzoprim (MBP), Cycloguanil (CYC), Trimetrexate (TMQ), Epiroprim (EPM), WR99210, and PDDF (CB3717)). The K_I values were obtained using an existing database, BindingDB) [27], and extensive literature searches. All inhibition constants in Table 1 had originally been generated in *in vitro* experiments using purified wild-type protein; all values were verified from the original publications. Much of the available inhibition data for ligand-DHFR pairs is published as IC_{50}s instead of K_I values. Ki and IC_{50} values for a ligand-enzyme couple can differ significantly (by 10-fold or even greater) depending on the experimental conditions and should not be compared to each other [28]. The drugs in this study are all single site competitive inhibitors allowing the use of the Prusoff equation to convert IC_{50}s to K_I values with the known substrate concentration used in the assay and the K_M for dihydrofolate for the appropriate DHFR homolog. We used the BotDB IC_{50} to K_I converter online server to verify the results of our calculations (Table 1) [29]. The K_M values used for the conversions are: 0.27 μM (*Ec*) [30], 0.50 μM (*Mt*) [31], 3.3 μM (*Tb*) [32], 0.70 μM (*Ma*) [33], 0.10 μM (*Hs*) [34], 0.60 μM (Tg) [35], 1.3 μM (*Lm*) [36], 23 μM (*Tc*) [32], 12.9 μM (*Pf*) [37], 107 μM (*Pv*) [38], 0.06 μM (*Rn*) [39], 2.7 μM (*Pc*) [40], and 1.0 μM (*Lc*) [41]. Data collection yielded a database containing 79 K_I values for 13 DHFR homologs. The ligands in Table 1 represent a wide variety of chemical structures rather than a set of very similar structures with minor functional group differences, enhancing the likelihood of identifying residues that are involved in ligand binding specificity in general, rather than ligand binding to a specific portion of the inhibitor structure.

2.2. Quantifying the similarity in inhibitor specificity for two DHFR homologs (S_{lig})

Once complete, the data in the table could have been converted to a yes/no relationship for inhibition using, for example, either $K_I = 0.1$ or 1 μM as the threshold for what is considered "to be a good inhibitor". This threshold would have simplified the quantification of the similarity in inhibitor specificity for two DHFR homologs. However, to avoid cases where the values straddled the cutoff value and where two relatively similar K_I values resulted in "different inhibition", we developed a novel approach to quantify the degree of similarity in inhibitor specificity profile for pairs of DHFR homologs. The S_{lig} values (ranging from 0 to 1.0; low to high similarity in inhibitor specificity) was determined using K_I data in Table 1. For this analysis, only homolog pairs that had data for at least two ligands in common were included. As a result, S_{lig} values were assigned to 77 of the 78 possible pairs of homologs between the selected 13 DHFRs. For each pair of DHFR homologs, S_{lig} was determined by the formula in Eq. (1). In this analysis, similar inhibition was defined as both homologs interacting *or* neither homolog interacting with a particular ligand.

$$S_{lig} = n / n_{total} \quad (1)$$

where

$n = \#$ of compounds that either inhibit both or neither homologs in a pair (criteria described)

$n_{total} = \#$ of compounds for which K_I data are available for both homologs in a pair

To determine n, rules for determining whether two homologs both inhibit a particular ligand were developed; two methods, the ratio method and the cutoff method, were used for the analysis. *Ratio method*: For each homolog pair, we determined a ratio *r* (smaller K_I value/larger K_I value) against each ligand (Fig. 1). If the ratio was between 1.0 and 0.01, then the two homologs were deemed to inhibit the ligand in a "similar" fashion (both ligands either inhibit or do not inhibit the homolog). If the ratio was smaller than 0.01, then the two homologs inhibit the ligand in a "dissimilar" fashion (one inhibits, the other does